LJ-01-01

SOUTHWEST FISHERIES SCIENCE CENTER

NATIONAL MARINE FISHERIES SERVICE - SOUTHWEST FISHERIES SCIENCE CENTER - P.O. BOX 271 LA JOLLA, CA 92038

June 26, 2001
2570 Dole Street

MARCH 2001

THE CETACEAN ZFX AND ZFY GENES: INTERFAMILIAL CHARACTERIZATION OF A NOVEL REGION FACILITATES GENDER DETERMINATION

by

Aviva Rosenberg and Sarah L. Mesnick

ADMINISTRATIVE REPORT LJ-01-01

"This report is used to ensure prompt dissemination of preliminary results, interim reports, and special studies to the scientific community. The material is not ready for formal publication since the paper may later be published in a modified form to include more recent information or research results. Abstracting, citing, or reproduction of this information is not allowed. Contact author if additional information is required."

THE CETACEAN ZFX AND ZFY GENES: INTERFAMILIAL CHARACTERIZATION OF A NOVEL REGION FACILITATES GENDER DETERMINATION

Aviva Rosenberg[1] and Sarah L. Mesnick[2]

[1]3444 Via Alicante
La Jolla, CA 92037
Tel: (858) 458 1928; FAX: (858) 458 5180
e-mail: rosenban@appliedbiosystems.com

[2]Southwest Fisheries Science Center
National Marine Fisheries Service/NOAA
P.O. Box 271
La Jolla, CA 92038
Tel: (858) 546-7148; FAX (858) 546-7003
e-mail: Sarah.Mesnick@noaa.gov

MARCH 2001

ADMINISTRATIVE REPORT LJ-01-01

ABSTRACT

Sequence polymorphism of homologues ZFX and ZFY, at a 604 base pair exon region, was examined in 10 known males and 10 known females across 7 cetacean families to determine the validity of a *Taq*1 restriction fragment length polymorphism (RFLP) assay for gender determination. Multiplex amplification, PCR cloning, and direct sequencing of these previously uncharacterized regions revealed: (1) 8 fixed differences between ZFX and ZFY, one of which was a *Taq1* restriction site at base pair 302 found only on the ZFX, (2) 29 variable sites between ZFX and ZFY that were not observed to form any *Taq*1 restriction sites, and (3) very low interspecific nucleotide diversity for both ZFX and ZFY across all families examined. These 3 findings support a ZFX/ZFY *Taq*1 RFLP method as a valid means to determine gender not only in the 10 cetacean species examined but most likely across the order Cetacea. Using this method, in conjunction with a test for the presence of Y-chromosomal gene SRY, provides the secondary confirmation necessary for positive sex identification. The use of proofreading enzyme *Pfu* DNA polymerase in PCR substantially improved data accuracy while a smaller amplicon size has enhanced the ability to gender samples from degraded DNA sources.

INTRODUCTION

Through deletion analysis of sex reversed individuals, homologous copies of the zinc finger protein genes, ZFX/ZFY, were mapped to interval 1A2 on the distal short arm of the X and Y chromosomes in humans (Page *et al.* 1987; Schneider-Gädicke *et al.* 1989). Additional studies excluded ZFY as the testis determining factor (TDF) (Palmer *et al.* 1989), the essential Y chromosomal gene necessary to induce testis formation from the bipotential gonad. Further analysis of XX males and XY females revealed the Y chromosomal gene SRY (Sinclair *et al.* 1990) and confirmed its role as the TDF (Berta *et al.* 1990; Jäger *et al.* 1990).

SRY detection via PCR (Saiki *et al.* 1988) has become a widely used gender determination method across a variety of mammalian species (Taberlet *et al.* 1993; Naito *et al.* 1994; Richard *et al.* 1994; Finch *et al.* 1996; Takahashi *et al.* 1998; Wilson & White 1998; S. R. Fain, J. P. LeMay, personal communication). Lack of SRY amplification, however, is not solely caused by gene absentia (Palsbøll *et al.* 1992; Taberlet *et al.* 1993) and may reflect lack of PCR amplification. To clarify the equivocal nature of negative results, some methods rely on an internal positive control (Taberlet *et al.* 1993; Finch *et al.* 1996; Takahashi *et al.* 1998; Wilson & White 1998; S. R. Fain, J. P. LeMay, personal communication) to indicate successful PCR amplification. Marine mammal specimens, for use in molecular ecology and population structure studies, are commonly obtained from three primary sources: (1) stranding events, where specimens are often at decomposition levels that prevent accurate assessment of reproductive organs, (2) forensic samples obtained from international markets, and (3) directed biopsy from live animals at sea. In such circumstances, additional confirmation of negative SRY assay results using an alternate, and preferably X chromosome, molecular based method is desired.

Because an X chromosomal specific marker responsible for ovarian development, the default pathway, has not yet been identified (Capel 1998), fixed differences between homologous copies of X and Y chromosomal genes, such as amelogenin (Nakahori *et al.* 1991) and ZFX/ZFY (Page *et al.* 1987; Schneider-Gädicke *et al.* 1989) have been used as alternate gender identification methods (Aasen & Medrano 1990; Nakahori *et al.* 1991; Palsbøll *et al.* 1992; Bérubé & Palsbøll 1996; Reynolds & Varlaro 1996).

The last exon of ZFX/ZFY, located most proximal to the centromere of four conserved segments and encoding 13 zinc-finger domains, is highly homologous among placental mammals, where ZFX has to be shown to escape X-inactivation (Page *et al.* 1987; Schneider-Gädicke *et al.* 1989; Jegalian & Page 1998). Despite this, restriction fragment length polymorphism (RFLP) analysis of region 96-1226 base pairs (bp) in cetaceans revealed an additional *Taq*1 site not found in humans (Palsbøll *et al.* 1992). Although this method did provide a gender specific restriction pattern for cetaceans, additional analysis of the last 700bp revealed that the fixed difference responsible for the results was not located in this latter portion of the exon. Hence, Bérubé & Palsbøll (1996) stated that the original assumptions of Palsbøll *et al.* (1992) were incorrect.

Although the *Taq*1 RFLP assay displays potential to be a viable method of gender determination in cetaceans, the validity of the method is uncertain. The essential portion of the exon has yet to be fully defined and characterized in cetaceans. The present study determines the ZFX and ZFY sequences from 96-699bp across 7 cetacean families, pinpointing the elusive *Taq*1 restriction sites, and also examines sequence polymorphisms across taxa and its effects, if any, on *Taq*1 restriction sites.

MATERIALS AND METHODS

Ten cetacean species from seven families were assayed. One known male and one known female from eight odontocete and two mysticete species were selected from the Southwest Fisheries Science Center (SWFSC) tissue archive (Table 1). Sex was determined by physical examination of the reproductive organs, except in the case of *P. macrocephalus* where determination was made by external examination of the genital slits.

Whole genomic DNA for all tissue samples was obtained using one of four extraction methods: (1) FastDNA kit (BIO101, Inc.), (2) hexadecyltrimethylammonium bromide (CTAB) (modified from Winnepenninckx *et al.* 1993), (3) 25:24:1 phenol/chloroform/isoamyl followed by ethanol precipitation (modified from Sambrook *et al.* 1989), or (4) lithium chloride (Gemmell & Akiyama 1996). Excluding *E.robustus* and the male *N. phocaenoides* samples, all tissue samples in this study were already extracted and stored frozen at –80°C. One microliter of the newly extracted DNA, along with 3µl (0.04µg/µl) High Range DNA size standard (BioRad), was electrophoresed on a 1% agarose minigel containing (0.15µg/ml) EtBr for 30 min at 85V in 1x TBE (pH 8.3), UV illuminated, and photographed. All twenty DNA extracts were quantitated on the TD700 Fluorometer (Turner Designs) using Picogreen dsDNA quantitation reagent (Molecular Probes).

Multiplex amplification of ZFX and ZFY, along with a negative control, was accomplished via a 50µl PCR (Saiki *et al.* 1988) reaction containing 5µl cloned *Pfu* DNA polymerase 10X Reaction Buffer (Stratagene), 0.2µM each of downstream primer ZFY0097 (Palsbøll *et al.* 1992) and upstream primer P2-3EZ (Aasen & Medrano 1990), 400µM dNTP, 1.25units cloned *Pfu* DNA polymerase (Stratagene), and 4-164ng genomic DNA. With a mineral oil overlay, cycling was performed on a Perkin Elmer Cetus DNA Thermal Cycler (TC1) with the following parameters: Hot start denaturation at 94°C for 2 min, followed by 30 cycles of denaturation at 94°C for 1 min, annealing at 48°C for 45 s, extension at 72°C for 1 min, followed by a final extension step at 72°C for 10 min. Five microliters of the PCR products, along with 0.5µl of 100bp Low DNA Ladder (Apex), were electrophoresed on 1% agarose minigels for 45 min as detailed above. The two amplicons produced are both 604bp and appear as one overlapping fragment on the minigel. Therefore, the following cloning procedure was use to obtain independent copies of ZFX and ZFY.

Pfu DNA polymerase amplicons do not possess non-templated 3' A overhangs. Therefore, PCR products were incubated with 2.5 units *Taq* DNA polymerase (Gibco-Brl Life Technologies) at 72°C for 10 min to generate the non-templated 3' A nucleotide addition. Following immediate clean up with Qiaquick PCR Purification Kit (Qiagen), 5µl of the clean PCR products were electrophoresed on 1% agarose minigels as above and quantitated on the TD700 Fluorometer using Picogreen. Using the TOPO TA Cloning® kit, clean PCR product was ligated with pCR®2.1-TOPO plasmid vector and transformed into TOP10™ One Shot Cells (Invitrogen). Two microliters of plasmid DNA, isolated with Ultraclean™ Mini Plasmid Prep Kit (Mo Bio Laboratories, Inc.), was electrophoresed on 1% agarose minigels for 30 min as above. One microliter of plasmid DNA was digested with 4 units *Eco*R1 (New England BioLabs) at 37°C for 1 h. Reactions were stopped with 2µl bromophenol blue Type IV buffer (Sambrook *et al.* 1989) and the full volume electrophoresed on 1% agarose minigels for 45 min as above to verify PCR product insertion.

Both strands were directly sequenced in a 15µl volume using 3µl of ABI Prism BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (PE Biosystems), 8µl of 1µM vector specific primers (TOPO TA Cloning® Instruction Manual) M13Reverse 5'-CAGGAAACAGCTATGAC-3' or T7 Promoter 5'-TAATACGACTCACTATAGGG-3', and 1- 4µl of plasmid DNA. Cycling parameters on a MJ Research PTC-100HB or Perkin Elmer 9600 Thermocycler were as follows: Hot start denaturation at 96°C for 5 min, followed by 30 cycles of denaturation at 96°C for 20 s, annealing at 52°C for 15 s, extension at 60°C for 4 min. Unincorporated dye-terminators were removed with an ethanol/sodium acetate precipitation modified from the BigDye™ Terminator Cycle Sequencing Ready Reaction Kit User Protocol. Sequences were run on an PE Biosystems ABI377 automated DNA sequencer using a 4.5% 29:1 polyacrylamide gel. Sequences were aligned, vector DNA excised, and edited with Sequencher v3.1.1 (Gene Codes Corporation).

Taq1 Restriction Digest

Ten microliters of each clean PCR product was digested in a mix of 1.5µl 10x *Taq*1 NE Buffer, 0.15µl BSA, 1µl of 1unit/µl *Taq*1 restriction endonuclease (New England BioLabs), and 2.35µl sterile water. Digest reactions were incubated at 65°C for 1 h and stopped by addition of 2µl bromophenol blue Type IV buffer. The complete digest reaction was electrophoresed on 1% agarose minigels with 0.15µg/ml EtBr for 1 h at 80V in 1xTBE.

RESULTS

A single allele from each female (XX) and two alleles from each male (XY) were detected from each of the 10 species. A majority-rule consensus sequence for all clones was generated for each of these alleles for all samples. The resultant 30 consensus sequences were subject to 1000 bootstrap replicates in a full heuristic search based on maximum parsimony using PAUP v.4.0b4a (Swofford 1999), with human ZFX and ZFY as outgroups. The search results produced two distinct clades (Fig. 1). The separation of cetacean from human alleles was supported by a confidence value of 99. All sequences attributed to females plus half the sequences attributed to males formed one clade, ZFX, supported by a confidence value of 99. The remaining 10 male consensus sequences formed the other clade, ZFY, supported by a confidence value of 96.

Excluding the non-templated A nucleotide addition at base 700, 604bp of ZFX/ZFY open reading frame (ORF) were sequenced (Fig. 2). Comparison of the ZFX allele obtained from males with the same allele obtained from females revealed 9 variable sites, of which sites 335 and 515

were only detected in the males of *K. breviceps* and *S. longirostris* hybrid, respectively (Fig. 3). Bérubé & Palsbøll (1996) did not examine specimens from the family Kogiidae resulting in the inability to detect variable site 650. A total of 8 different ZFX haplotypes were observed in the 20 sequences. All the ZFX variable sites displayed synonymous substitutions resulting in a lack of amino acid heterozygosity across all the families examined (Fig. 4).

The 10 ZFY sequences displayed 22 variable sites resulting in 8 different haplotypes (Fig. 3). Seven of the nucleotide variable sites produced 6 amino acid variable positions resulting in 5 amino acid haplotypes (Fig. 3 and Fig. 4), all of which maintain the ORF. Kogiidae exhibited a 3bp deletion at 379-381bp on the ZFY causing a deletion of amino acids 95 and 96 that did not affect the ORF.

Comparison of ZFX and ZFY from all species revealed 8 fixed differences, all synonymous, and 29 variable sites, of which 2 were variable on both ZFX and ZFY (Fig. 3). The fixed difference at position 302 forms a *Taq*1 restriction site in ZFX but not in ZFY. An additional *Taq*1 restriction location (denoted by the TCGA sequence) was found for both ZFX and ZFY beginning at position 144. No variability was detected at any of the restriction sites. The restriction pattern for a ZFX fragment exhibits sizes 49, 157, and 399. Likewise, a ZFY fragment exhibits sizes 49 and 556. Therefore, females (XX) display the fragments 49, 157, and 399 whereas males (XY) display the fragments 49, 157, 399, and 556 (Fig. 5). All fragments are displayed in equal intensity on a minigel.

The high haplotypic diversity observed in both ZFX and ZFY results from very low nucleotide diversity. A pairwise uncorrected "p" distance matrix of all ZFX haplotypes estimated nucleotide diversity between 0.2%-1.2% with a mean equal to 0.6%. A similar calculation for all ZFY haplotypes estimated nucleotide diversity between 0.2%- 2.3% with a mean equal to 1.4%. The above two calculations were repeated with the inclusion of the human ZFX and human ZFY, respectively. The human ZFX was different from the cetacean ZFX by 3.3%-4.0% with a mean equal to 3.6%. The human ZFY was different from the cetacean ZFY by 5.5-6.3% with a mean equal to 6.0%. All distance calculations were computed in PAUP v.4.0b4a (Swofford 1999).

Anomalous Data

A total of 22 sequences of ZFX and ZFY exhibited replication errors. Approximately 1/3 of the anomalous sequences were resequenced providing results concordant with the original data. The majority of the anomalous sequences had only a single point mutation. A total of 26 sites, spread throughout the 605bp, showed replication errors. The majority of these sites displayed an error in one sequence. Sites displaying an error in more than one sequence exhibited consistency across sequences, except in one instance, regarding the anomalous change. Twenty-seven transitions, 10 transversions, 1 insertion in a $(A)_n$ tract, 2 deletions in a different $(A)_n$ tract, and a recombinant of ZFX/ZFY were observed in the anomalous clones, of which some led to stop codon(s) creation and interruption of the ORF. All anomalous sequences excluding the recombinant, in conjunction with the 30 consensus sequences, were subject to a heuristic search with 100 bootstrap replicates similar to the one noted earlier (figure not shown). All the anomalous sequences grouped as expected in either ZFX or ZFY. A separate similar search using 500 bootstrap replicates was conducted for the ZFX/ZFY recombinant. The recombinant formed a clade with the ZFY sequences supported by a confidence value of 62 (figure not shown). This recombinant sequence was therefore excluded from any additional analyses.

DISCUSSION

Sample Selection

To determine with confidence the X-chromosomal and Y-chromosomal sequences of the zinc finger protein gene in cetaceans, known males and known females are critical. In non-dimorphic cetacean species, examination of reproductive organs or external examination of the genital slit in relation to the anal slit (Jefferson *et al.* 1994) are the only two methods that can provide physical evidence of a specimen's sex, with the former providing an even higher level of confidence. The SWFSC tissue archive contained 19 odontocete and 2 mysticete species that had at least 1 male and 1 female for which reproductive organ data had been recorded. Although this study aimed to represent the 14 extant cetacean families, paucity of gender information prevented the inclusion of the families Neobalaenidae, Balaenopteridae, Ziphiidae, Plantanistidae, Iniidae, Lipotidae, and Pontoporiidae. Although specimens were available for an additional 1 phoceonid and 11 delphinid species, these species were not examined. Because lack of variability was detected across interfamilial boundaries at *Taq*1 restriction sites on the ten species examined in this study, additional intrafamilial data collection, from these particular families, was assumed unnecessary in testing for the validity of the method.

Recent human studies have demonstrated the ability to detect male fetal cells, using PCR amplification of the ZFY and/or SRY genes, not only in the maternal circulation (Liou *et al.* 1993; Lo *et al.* 1999) but also in skin lesions of pregnant females (Aractingi *et al.* 1998). As this phenomenon has yet to be studied in cetaceans and to avoid the possibility of amplifying male fetal DNA instead of maternal DNA, resulting in possible errors in ZFX allele identification, pregnant females were avoided in sample selection. Although the female sperm whale was not necropsied to prove lack of pregnancy, genetic results indicate that the specimen was either not pregnant, was carrying a female fetus, not carrying a male fetus, or the Y copy of the male fetal DNA not detected. Amplification of female fetal DNA would not have interfered with identification of the ZFX allele.

Data accuracy and verification

Possessing 3'⇒5' exonuclease-dependent proofreading activity, *Pfu* DNA polymerase minimizes PCR amplification errors by greater than ten-fold compared to *Taq* DNA polymerase (Lundberg *et al.* 1991), the enzyme utilized in Palsbøll *et al.* (1992). *Pfu* DNA polymerase, possessing a lower error rate (1.3×10^{-6}) than any other proofreading enzyme available (Cline *et al.* 1996), proved to be essential for the high-fidelity techniques of PCR cloning and RFLP analysis of ZFX/ZFY. In addition, because the *Taq*1 RFLP assay was found to be dependent on a single nucleotide difference, the necessity of a proofreading enzyme was further warranted.

The P2-3EZ primer sequence was originally confirmed as an exact match to cetaceans using the sequence data presented in Bérubé & Palsbøll (1996). Data obtained herein were fully consistent with overlapping sequence data of regions 586-606bp (primer ZFYX0582F) and 622-674bp in Bérubé & Palsbøll (1996). It should be noted that the Fig.2 sequence of Bérubé & Palsbøll (1996) was mistakenly misnumbered and should read as 586-1226bp (M. Bérubé & P. Palsbøll, personal communication).

Concordance of sequence data obtained from two independent rounds of PCR and cloning executed on 5 of the 20 samples also served to confirm data accuracy (Table 1). The consensus sequence for 3 other samples was determined from only 2 individual sequences but were produced

from a single PCR: BLm ZFY, SWm ZFX and PSm ZFX. BLm ZFY exhibited a unique haplotype, as similarly seen in BWm ZFY, but neither contributed any additional ZFY variable sites. SWm ZFX exhibited the same haplotype as SWf ZFX, a trend observed in the majority of the cetacean species examined. One of the 2 sequences obtained for PSm ZFX contained a nonsynonymous base substitution at a nonvariable site. Because all the other ZFX sequences were observed to lack amino acid variability, the base substitution was regarded as a replication error. In addition, ZFX of PSm was identical to PSf ZFX.

Although template DNA is a potential source for the observed mutations in the anomalous sequences presented in the results, the frequency exhibited exceeds that expected for *Pfu* amplified products as referenced above. The TOP10™ *E.coli* strain used in the cloning procedure possesses the MutHSL post-replicative mismatch repair system, one of the most extensively studied in a living organism (Marra & Schär 1999). Fujii *et al.* (1999), using the rpsL gene, compared mutation assay results of an in vitro system lacking DNA repair enzymes, a similar in vivo system using mismatch repair deficient cells, and a wild-type *E.coli* strain. Their results demonstrated that replication errors are produced by the replicative apparatus of *E.coli*. Although the repair efficiency of the wild-type *E.coli* strain is much greater than that in deficient systems, single-base frameshifts at runs of the same nucleotide, possibly due to melting-misalignment, and base mismatches are generated that remain uncorrected. Proofreading error of the DNA polymerase III holoenzyme of the replicative apparatus was proposed as the progenitor of these mutations in vitro and in vivo. In addition, the wild-type strain mismatch repair system was found to correct single base frameshifts much more efficiently than transitions or transversions. These findings suggest that such mutations are likely due to the replication apparatus of *E.coli*.

RFLP analysis

Two major improvements made in the *Taq*1 RFLP assay, use of proofreading enzyme *Pfu* DNA polymerase and halving the size of ZFX/ZFY amplicons presented in Palsbøll *et al.* (1992), have served to enhance both data reliability and the possibility to determine gender from degraded DNA. Proper and efficient execution of the assay, however, is dependent upon quantitation of DNA extracts prior to PCR and, more importantly, of PCR products prior to the RFLP assay. As demonstrated (Fig. 5), using less than 20ng of amplicon inhibits product visibility while digest reactions containing in excess of 120ng will likely fail to complete digestion under the stated conditions as seen in lane 8 of Fig. 5. Ensuring complete digestion of the amplicon is critical to avoid adversely assigning a faint 605bp fragment (the incompletely digested original amplicon) as a 556bp ZFY restriction fragment, which would result in an erroneous sex determination.

Combined with the sequence data from Fig. 2 of Bérubé & Palsbøll (1996), the 1131bp amplicon from 96-1226bp presented in Palsbøll *et al.* (1992) contains 3 *Taq*1 restriction sites on ZFX located at 144bp, 301bp, and 765bp, and 2 *Taq*1 restriction sites on ZFY located at 144bp and 765bp. Using their primers, the restriction pattern for a ZFX would exhibit fragment sizes: 49, 157, 464, and 461. ZFY would exhibit fragments sizes: 49, 621, and 461. Females (XX) would display the fragments 49, 157, 464, and 461 whereas males (XY) would display the fragments 49, 157, 464, 461, and 621. Due to their combined fluorescence, overlapping fragments 464 and 461 may possibly appear brighter in comparison to the other fragments, especially in males. Lastly, the assumed fragment sizes published in Palsbøll *et al.* (1992) for ZFX (182, 439) and ZFY (439, 621) are remarkably similar to the true fragment sizes alleviating some concerns regarding possible misidentifications when using the latter protocol.

Limitations

Following an interspecific comparison among marine mammals, both the ZFX and ZFY genes were found to be highly conserved at both the nucleotide and amino acid levels. Although a high degree of conservation is necessary in order to maintain genetic functionality, this study demonstrates that differences at the nucleotide and amino acid level in ZFX and ZFY do exist among marine mammals. A comparison of ZFX and ZFY from humans and cetaceans revealed that these differences could also affect the presence or absence of *Taq*1 restriction site locations. The gender specific restriction patterns evident for these taxa are the result of different *Taq*1 restriction site locations. High homology in ZFX/ZFY among placental mammals, therefore, does not guarantee equivalence in data. It is therefore recommended that the sequences provided herein should not be applied for gender determination in different mammalian species other than those belonging to the order Cetacea without further study.

CONCLUSIONS

This ZFX/ZFY *Taq*1 RFLP method has been found in this study to be a valid means to determine gender not only in the 10 cetacean species examined but most likely across the order Cetacea. This method can provide the secondary confirmation necessary for positive sex identification in marine mammal specimens. Using the proofreading enzyme *Pfu* DNA polymerase and halving the size of ZFX/ZFY amplicons have enhanced the *Taq*1 RFLP assay. As the Taq1 restriction site locations have been validated in this study, further developments in cetacean gender determination can focus attention on testing available technologies that may increase the speed and/or efficiency for screening samples and interrogating this SNP (single nucleotide polymorphism), e.g., Snapshot™ ddNTP Primer Extension Kit (Applied Biosystems Foster City, CA).

ACKNOWLEDGEMENTS

Funding for this study was provided by the National Marine Fisheries Service. Many thanks to the following people (in no particular order) who donated samples to the Southwest Fisheries Science Center tissue archive making this study possible: Pierre-Yves Doust (University of Prince Edward Island), Robert Suydam (North Slope Borough, Barrow, AK), Diane Paschall (The Marine Mammal Center, Sausalito, CA), Daniel K. Odell (Sea World, FL), John Heyning and Dave Janiger (LA County Natural History Museum), Graham Worthy (Texas A&M University), Tom Jefferson (Ocean Park, Hong Kong), Deborah Thiele (Wildlife Management Environment Australia) and Debbie Duffield (Portland State University). A special thanks to Andrew Dizon for wholeheartedly supporting this research, Kelly Robertson for providing technical assistance during sample selection, Rick LeDuc for assistance in data analysis, Greg O'Corry-Crowe for providing A. Rosenberg's first glimpse at gender identification, our reviewers, and gratefully Per Palsbøll and Martine Bérubé (University of Wales, Bangor, UK) for their continued support and cooperation throughout this project.

LITERATURE CITED

Aasen E, Medrano JF (1990) Amplification of the ZFY and ZFX genes for sex identification in humans, cattle, sheep, and goats. *Bio/Technology*, **8**, 1279-1281.

Aractingi S, Berkane N, Bertheau P *et al.* (1998) Fetal DNA in skin of polymorphic eruptions of pregnancy. *Lancet*, **352**, 1898-1901.

Berta P, Hawkins JR, Sinclair AH *et al.* (1990) Genetic evidence equating SRY and the testis-determining factor. *Nature*, **348**, 448-450.

Bérubé M, Palsbøll P (1996) Identification of sex in Cetaceans by multiplexing with three ZFX and ZFY specific primers. *Molecular Ecology*, **5**, 283-287.

Bérubé M, Palsbøll P (1996) Erratum for Identification of sex in Cetaceans by multiplexing with three ZFX and ZFY specific primers. *Molecular Ecology*, **5**, 602.

Capel B (1998) Sex in the 90s: SRY and the switch to the male pathway. *Annual Review of Physiology*, **60**, 497-523.

Cline J, Braman JC, Hogrefe HH (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research*, **24**, 3546-3551.

Finch JL, Hope RM, van Daal A (1996) Human sex determination using multiplex polymerase chain reaction (PCR). *Science & Justice*, **36**, 93-95.

Fujii S, Akiyama M, Aoki K *et al.* (1999) DNA replication errors produced by the replicative apparatus of *Escherichia coli*. *Journal of Molecular Biology*, **289**, 835-850.

Gemmell NJ, Akiyama S (1996) An efficient method for the extraction of DNA from vertebrate tissues. *Trends in Genetics*, **12**, 338-339.

Jäger RJ, Anvret M, Hall K, Scherer G (1990) A human XY female with a frame shift mutation in the candidate testis-determining gene SRY. *Nature*, **348**, 452-453.

Jefferson TA, Myrick, Jr. AC, Chivers SJ (1994) *Small Cetacean Dissection and Sampling: A Field Guide* NOAA-TM-NMFS-SWFSC-198.

Jegalian K, Page DC (1998) A proposed path by which genes common to mammalian X and Y chromosomes evolve to become X inactivated. *Nature*, **394**, 776-780.

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park, PA.

Liou JD, Pao CC, Hor JJ, Kao SM (1993) Fetal cells in the maternal circulation during first trimester in pregnancies. *Human Genetics*, **92**, 309-311.

Lo YMD, Leung TN, Tein MSC *et al.* (1999) Quantitative abnormalities of fetal DNA in maternal serum in preeclampsia. *Clinical Chemistry*, **45**, 184-188.

Lundberg KS, Shoemaker DD, Adams MWW *et al.* (1991) High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene*, **108**, 1-6.

Marra G, Schär P (1999) Recognition of DNA alterations by the mismatch repair system. *Journal of Biochemistry*, **338**, 1-13.

Nakahori Y, Takenaka O, Nakagome Y (1991) A human X-Y homologous region encodes "Amelogenin". *Genomics*, **9**, 264-269.

Nakahori Y, Hamano K, Iwaya M, Nakagome Y (1991) Sex identification by polymerase chain reaction using X-Y homologous primer. *American Journal of Medical Genetics*, **39**, 472-473.

Page DC, Mosher R, Simpson EM *et al.* (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell*, **51**, 1091-1104.

Palmer MS, Sinclair AH, Berta P *et al.* (1989) Genetic evidence that ZFY is not the testis-determining factor. *Nature*, **342**, 937-939.

Palsbøll PJ, Vader A, Bakke I, El-Gewely MR (1992) Determination of gender in cetaceans by the polymerase chain reaction. *Canadian Journal of Zoology*, **70**, 2166-2170.

Reynolds R and Varlaro J (1996) Gender determination of forensic samples using PCR amplification of ZFX/ZFY gene sequences. *Journal of Forensic Sciences*, **41**, 279-286.

Rice DW (1998) Marine mammals of the world: systematics and distribution. *Special Publication Number 4 of the Society for Marine Mammalogy*, **4**, 51-126.

Richard KR, McCarrey SW, Wright JM (1994) DNA sequence from the SRY gene of the sperm whale (*Physeter macrocephalus*) for use in molecular sexing. *Canadian Journal of Zoology*, **72**, 873-877.

Saiki RK, Gelfand DH, Stoffel S *et al.* (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*, **239**, 487-491.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual,* 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Schneider-Gädicke A, Beer-Romero P, Brown LG, Nussbaum R, Page DC (1989) ZFX has a gene structure similar to ZFY, the putative human sex determinant, and escapes x inactivation. *Cell*, **57**, 1247-1258.

Sinclair AH, Berta P, Palmer MS *et al.* (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, **346**, 240-244.

Swofford DL (1999) PAUP. Phylogenetic Analysis Using Parsimony (and Other Methods). Version v.4.0b4a Sinauer Associates, Sunderland, Massachusetts.

Taberlet P, Mattock H, Dubois-Paganon C, Bouvet J (1993) Sexing free-ranging brown bears *Ursus arctos* using hairs found in the field. *Molecular Ecology*, **2**, 399-403.

Takahashi M, Masuda R, Uno H *et al.* (1998) Sexing of carcass remains of the Sika deer (*Cervus nippon*) using PCR amplification of the SRY gene. *Journal of Veterinary Medical Science*, **60**, 713-716.

Wilson PJ, White BN (1998) Sex identification of elk (*Cervus elaphus canadensis*), moose (*Alces alces*), and white-tailed deer (*Odocoileus virginianus*) using the polymerase chain reaction. *Journal of Forensic Science*, **43**, 477-482.

Winnepenninckx B, Backeljau T, DeWachter R (1993) Extraction of high molecular weight DNA from mollusks. *Trends in Genetics*, **9**, 407.

Table 1. Eight odontocete and two mysticete species examined from seven families, code abbreviations, and Genbank accession numbers.

Family	Species	Common Name	Code	Genbank Accession No.		
				ZFX (f)	ZFX (m)	ZFY
Balaenidae	*Balaena mysticetus*	bowhead whale	BW	AF260783	AF260784	AF260785
Eschrichtiidae	*Eschrichtius robustus*	gray whale	GW	AF260789	AF260790	AF260791
Physeteridae	*Physeter macrocephalus*	sperm whale	SW	AF260801	AF260802	AF260803
Kogiidae	*Kogia breviceps*	pygmy sperm whale	PS	AF260792	AF260793	AF260794
	Kogia sima	dwarf sperm whale	DS	AF260797	AF260795	AF260796
Monodontidae	*Delphinapterus leucas*	Beluga	BL	AF260788	AF260786	AF260787
Delphinidae	*Stenella longirostris orientalis*	eastern spinner dolphin	ES	AF260812	AF260810	AF260811
	Stenella longirostris hybrid	whitebelly spinner dolphin*	WS	AF260807	AF260808	AF260809
Phocoenidae	*Phocoena phocoena*	harbor porpoise	HP	AF260804	AF260805	AF260806
	Neophocaena phocaenoides	finless porpoise	FP	AF260800	AF260798	AF260799

*Intergrade form or hybrid between *Stenella longirostris longirostris* and *Stenella longirostris orientalis*. The subspecies status is not yet defined (Rice 1998).

Fig. 1 50% majority-rule consensus tree of the 30 cetacean sequences following 1000 bootstrap replicates and rooted with outgroups HuZFX (human ZFX; Schneider-Gädicke *et al.* 1989) and HuZFY (human ZFY; Page *et al.* 1987). f=female. m=male. 1=allele X. 2=allele Y. All characters are unordered and of equal weight. The starting tree was obtained via stepwise addition. Sequence addition followed the "simple" algorithm using BWf as the reference taxon. Steepest descent and the TBR branch swapping algorithm were in effect.

Fig. 2 ZFX/ZFY nucleotide and amino acid consensus sequences from the bowhead whale (BW) compared to human ZFX/ZFY. Site numbering according to human ZFX (Schneider- Gädicke *et al.* 1989) and human ZFY (Page *et al.* 1987) sequences. Dashes denote nucleotide agreement with BW ZFX. Gray highlighted amino acids indicate variable positions on the cetacean ZFY. Primer regions are underlined and names noted above. *Taq*1 restriction sites are double underlined. The amino acid sequence from bowhead whales is identical on the ZFX and ZFY. A comparison of all cetacean ZFY sequences to the human ZFY revealed 27 fixed differences (not shown). A comparison of all cetacean ZFX sequences to the human ZFX revealed 18 fixed differences (not shown).

Fig. 3 Comparison of ZFX and ZFY fixed and variable sites. f=female m=male. Number in parentheses equals the number of clones that displayed replication errors. Gray highlighted positions denote 8 fixed differences between ZFX and ZFY. Arrows indicate sites variable on both ZFX and ZFY. ZFY variable sites 208, 251, 256, 336, 337, 567, and 664 display nonsynonymous substitutions.

Fig. 4 Amino acid variable site comparison between cetacean ZFX/ZFY and human ZFX/ZFY. Gray highlighted site indicates a fixed amino acid difference between cetaceans and humans on both the ZFX and ZFY. Sites 145 and 176 display fixed differences between the cetacean ZFY and human ZFY.

Fig. 5 *Taq*1 restriction digest minigels of *Pfu* amplified PCR products. From left to right, (A) lane 1, SWf; lane 2, SWm; lane 3, PSf; lane 4, PSm; lane 5 and 6, DSf; lane 7, DSm; lane 8, BLf (incomplete digestion resulted in a faint PCR product remnant; 274ng was utilized in the digest reaction); lane 9, BLm; lane 10, HPf; lane 11, HPm; lane 12 and 13, FPf; lane 14 and 15, FPm; lane 16, WSf; lane 17, WSm; lane 18, ESf; lane 19, ESm; lane 20, 100bp ladder with fragments 100-1000bp. Lane 6 utilized only 8ng of PCR product in the restriction assay (versus 45ng in lane 5), lanes 12 and 13 both had less than 10ng, and lane 14 only had 2ng (versus 17ng in lane 15). (B) lane 1=BWf; lane 2 and 3=BWm; lane 4=GWf; lane 5 and 6= GWm; lane 7= stranded sperm whale unknown (displays incomplete digestion for the same reasons as BLf); lane 8=100bp ladder. Lane 4 had less than 1 ng, and lane 5 only had 3ng (versus 17ng in lane 6). An adequate amount of each allele for samples DSf, FPf, FPm, BWm, and GWm were not obtained during the initial cloning procedure resulting in subjugation of these samples to an additional PCR. The 49bp band is not visible in most samples, possibly due to the small fragment size. The unknown stranded specimen was determined to be a female via PCR cloning of the 605bp amplification product and negative SRY assay (Richard *et al.* 1994) results. This unknown was restricted a second time using less PCR product (40ng versus the original 168ng) resulting in complete digestion (photo not shown).

Figure 1

Bootstrap

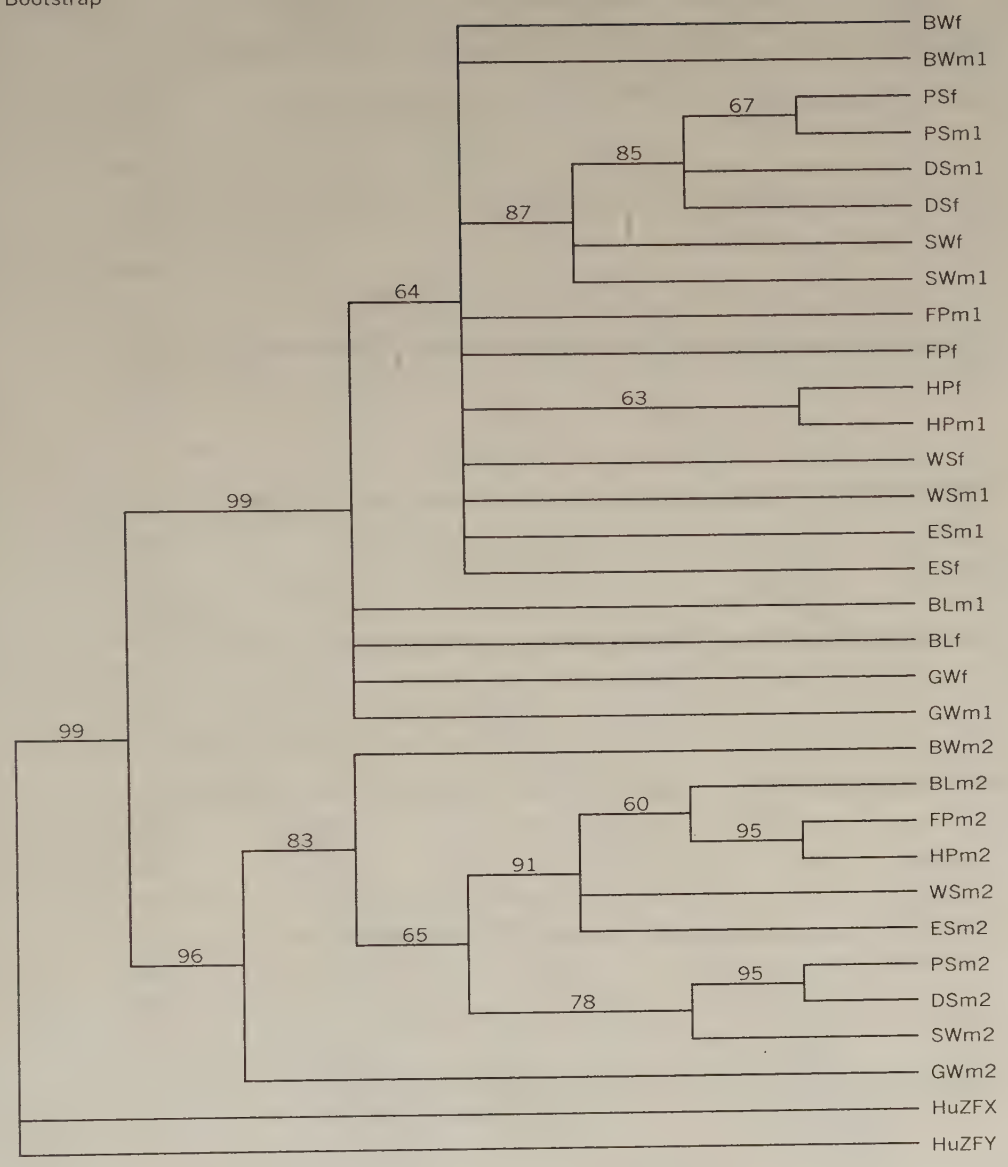


Fig. 2

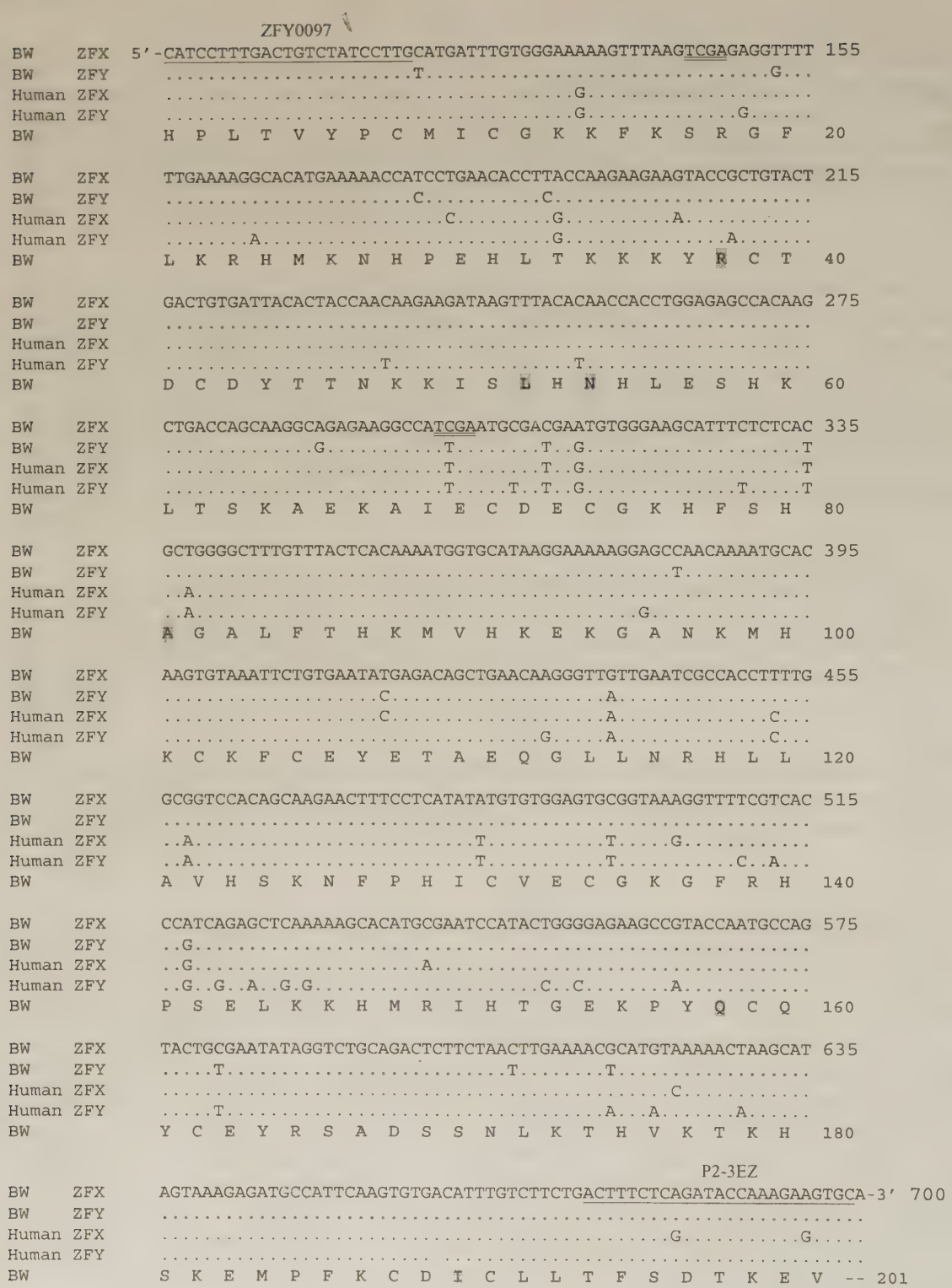

ZFY0097
BW ZFX 5'-CATCCTTTGACTGTCTATCCTTGCATGATTTGTGGGAAAAAGTTTAAGTCGAGAGGTTTT 155
BW ZFY .....T.....G...
Human ZFX .....G.....
Human ZFY .....G.....G.....
BW H P L T V Y P C M I C G K K F K S R G F 20
BW ZFX TTGAAAAGGCACATGAAAAACCATCCTGAACACCTTACCAAGAAGAAGTACCGCTGTACT 215
BW ZFY .....C.....C.....
Human ZFX .....C.....G.....A.....
Human ZFY .....A.....G.....A.....
BW L K R H M K N H P E H L T K K K Y R C T 40
BW ZFX GACTGTGATTACACTACCAACAAGAAGATAAGTTTACACAACCACCTGGAGAGCCACAAG 275
BW ZFY .....
Human ZFX .....
Human ZFY .....T.....T.....
BW D C D Y T T N K K I S L H N H L E S H K 60
BW ZFX CTGACCAGCAAGGCAGAGAAGGCCATCGAATGCGACGAATGTGGGAAGCATTTCTCTCAC 335
BW ZFY .....G.....T.....T..G.....T
Human ZFX .....T.....T..G.....T
Human ZFY .....T.....T..T..G.....T.....T
BW L T S K A E K A I E C D E C G K H F S H 80
BW ZFX GCTGGGGCTTTGTTTACTCACAAAATGGTGCATAAGGAAAAAGGAGCCAACAAAATGCAC 395
BW ZFY .....T.....
Human ZFX ..A.....
Human ZFY ..A.....G.....
BW A G A L F T H K M V H K E K G A N K M H 100
BW ZFX AAGTGTAAATTCTGTGAATATGAGACAGCTGAACAAGGGTTGTTGAATCGCCACCTTTTG 455
BW ZFY .....C.....A.....
Human ZFX .....C.....A.....C...
Human ZFY .....G.....A.....C...
BW K C K F C E Y E T A E Q G L L N R H L L 120
BW ZFX GCGGTCCACAGCAAGAACTTTCCTCATATATGTGTGGAGTGCGGTAAAGGTTTTCGTCAC 515
BW ZFY .....
Human ZFX ..A.....T.....T.....G.....
Human ZFX ..A.....T.....T.....C..A...
BW A V H S K N F P H I C V E C G K G F R H 140
BW ZFX CCATCAGAGCTCAAAAAGCACATGCGAATCCATACTGGGGAGAAGCCGTACCAATGCCAG 575
BW ZFY ..G.....
Human ZFX ..G.....A.....
Human ZFX ..G..G..A..G.G.....C..C.....A.....
BW P S E L K K H M R I H T G E K P Y Q C Q 160
BW ZFX TACTGCGAATATAGGTCTGCAGACTCTTCTAACTTGAAAACGCATGTAAAAACTAAGCAT 635
BW ZFY .....T.....T.....T.....
Human ZFX .....C.....
Human ZFX .....T.....A...A.....A.....
BW Y C E Y R S A D S S N L K T H V K T K H 180
P2-3EZ
BW ZFX AGTAAAGAGATGCCATTCAAGTGTGACATTTGTCTTCTGACTTTCTCAGATACCAAAGAAGTGCA-3' 700
BW ZFY .....
Human ZFX .....G.....G.....
Human ZFX .....
BW S K E M P F K C D I C L L T F S D T K E V -- 201

Fig. 3

	1111122222 2333333344 4444445555 5556666	Number of Clones
	1579900355 9011333801 3456891145 6680156	
	9291738616 0214567346 7681875887 3718704	
ZFX		
BWf	CTTTGGGCAA ACCACGCCAT GCGCTCCATG GCCCGAT	4 (2)
BWm	....	10
GWf	....T.....	10
GWm	....T.....	9
SWf	.....A.... ....A	13
SWm	.....A.... ....A	2
PSf	.....A.... .....G. ....C....A .....G.	9 (1)
PSm	.....A.... ....T...G. ....C....A .....G.	2 (1)
DSf	.....A.... ....C....A .....G.	9 (2)
DSm	.....A.... ....C....A .....G.	4
BLf	....T.....	8 (1)
BLm	....T.....	11 (2)
ESf	....	9 (1)
ESm	....	4
WSf	....	7
WSm	....T...	9 (1)
HPf	.... .T.....	5
HPm	.... .T.....	9
FPf	....	3 (1)
FPm	....	14
ZFY		
BWm	TGCC..... GTTGT..T.C A.....G.. ..TTT..	2
GWm	T.CC..... .TTG...T.. A.A....GC. .GTTT..	8
SWm	TCCC..A... GTTG.AGT.C A.A.C..G.. ..TTC..	6 (2)
PSm	TCCCA.A..G GTTG...T.C A.A.C..... ..TTT..	6 (1)
DSm	TCCCA.A..G GTTG...T.C A.A.C..... ..TTT..	9 (2)
BLm	T.CC..A.T. GTTG...T.C ..AACT.G.. ..TTT.G	2
ESm	T.CC..A.T. GTTG...T.C A.AA.T.G.. ..TTT.G	8 (2)
WSm	T.CC..A.T. GTTG...T.C A.AA.T.G.. ..TTT.G	5
HPm	T.CC..ATT. .TTGT..T.C ..AA.T.G.. A.TTT.G	5 (1)
FPm	T.CC..ATT. .TTG...T.C ..AA.T.G.. A.TTT.G	10 (1)

Fig. 4

```
                      1111
                    3355845 79
                    3824158 60
BWm ZFY             TRLNAKQ VI
GWm ZFY             ......E ..
SWm ZFY             .H..S.. ..
PSm ZFY             .H.S... ..
DSm ZFY             .H.S... ..
BLm ZFY             .HF.... .S
ESm ZFY             .HF.... .S
WSm ZFY             .HF.... .S
HPm ZFY             .HF.... .S
FPm ZFY             .HF.... .S
Human ZFY           AH...R. I.
Human ZFX           A..... ..
Cetacean ZFX        ..... ..
```

Fig. 5A

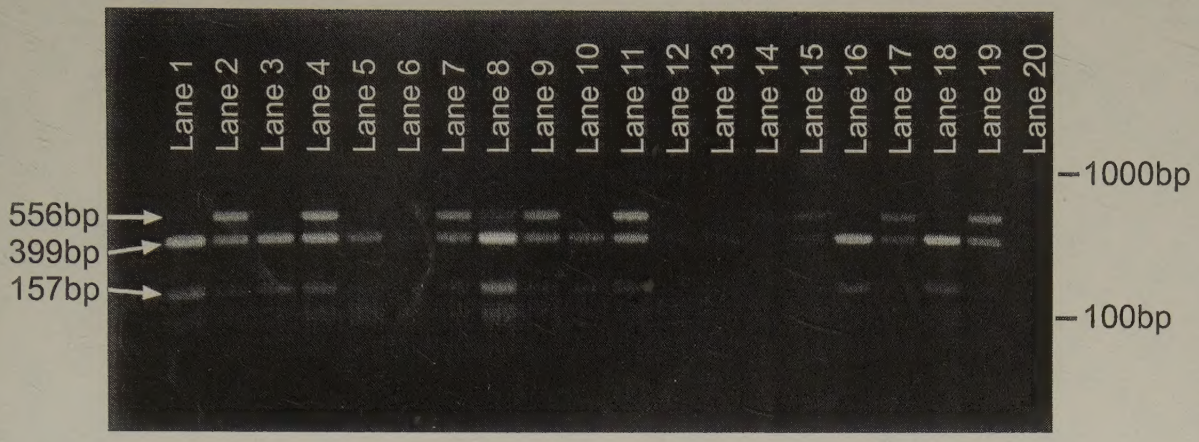
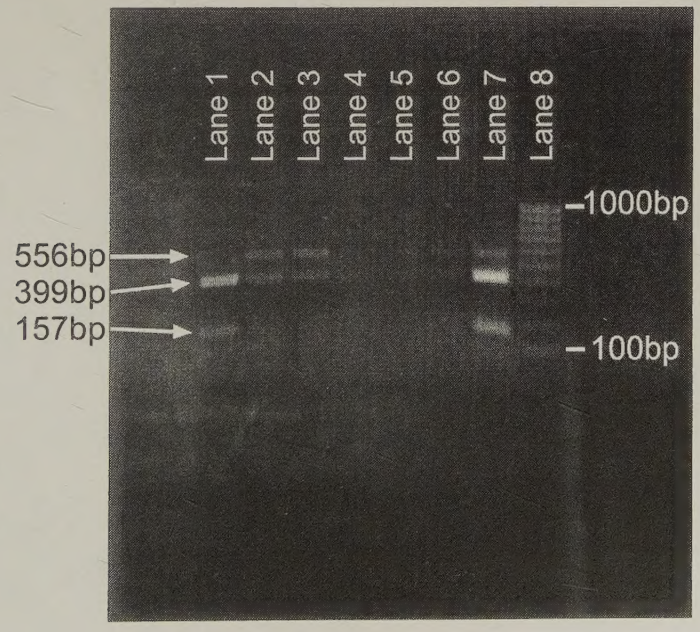
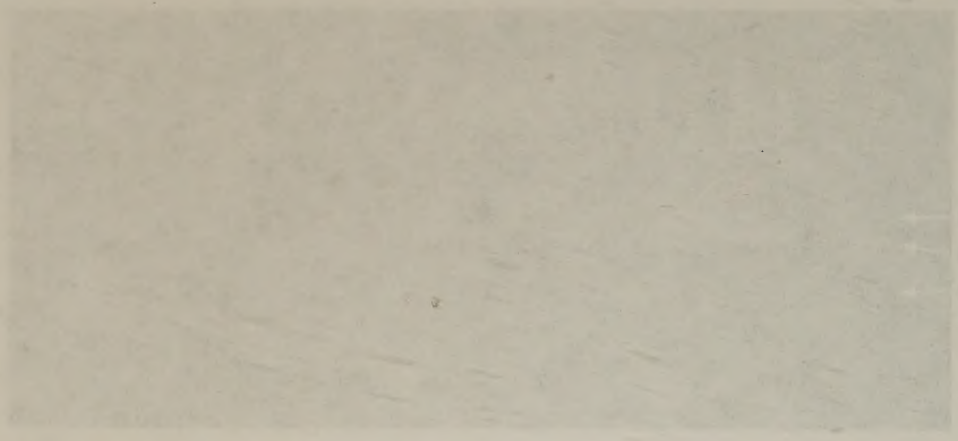


Fig. 5 *Taq* 1 restriction digest minigels of *Pfu* amplified PCR products. From left to right, (A) lane 1, SWf; lane 2, SWm; lane 3, PSf; lane 4, PSm; lane 5 and 6, DSf; lane 7, DSm; lane 8, BLf (incomplete digestion resulted in a faint PCR product remnant; 274ng was utilized in the digest reaction); lane 9, BLm; lane 10, HPf; lane 11, HPm; lane 12 and 13, FPf; lane 14 and 15, FPm; lane 16, WSf; lane 17, WSm; lane 18, ESf; lane 19, ESm; lane 20, 100bp ladder with fragments 100-1000bp. Lane 6 utilized only 8ng of PCR product in the restriction assay (versus 45ng in lane 5), lanes 12 and 13 both had less than 10ng, and lane 14 only had 2ng (versus 17ng in lane 15). (B) lane 1=BWf; lane 2 and 3=BWm; lane 4=GWf; lane 5 and 6= GWm; lane 7= stranded sperm whale unknown (displays incomplete digestion for the same reasons as BLf); lane 8=100bp ladder. Lane 4 had less than 1 ng, and lane 5 only had 3ng (versus 17ng in lane 6). An adequate amount of each allele for samples DSf, FPf, FPm, BWm, and GWm were not obtained during the initial cloning procedure resulting in subjugation of these samples to an additional PCR. The 49bp band is not visible in most samples, possibly due to the small fragment size. The unknown stranded specimen was determined to be a female via PCR cloning of the 605bp amplification product and negative SRY assay (Richard *et al.* 1994) results. This unknown was restricted a second time using less PCR product (40ng versus the original 168ng) resulting in complete digestion (photo not shown).

Fig. 5B





RECEIVED
JUN 18 2001
NMFS/HONOLULU LAB
DIRECTORS OFFICE